



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> <b>C12Q 1/68</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/65094</b> <b>(43) International Publication Date:</b> 2 November 2000 (02.11.00)
<b>(21) International Application Number:</b> PCT/US00/10879 <b>(22) International Filing Date:</b> 21 April 2000 (21.04.00) <b>(30) Priority Data:</b> 60/130,563                      22 April 1999 (22.04.99)                      US <b>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application</b> US    60/130,563 (CIP) Filed on    22 April 1999 (22.04.99) <b>(71) Applicant (for all designated States except US):</b> THE ALBERT EINSTEIN COLLEGE OF MEDICINE OF YESHIVA UNIVERSITY [US/US]; 1300 Morris Park Avenue, Bronx, NY 10461 (US). <b>(72) Inventor; and</b> <b>(75) Inventor/Applicant (for US only):</b> SINGER, Robert, H. [US/US]; 144 East 84th Street, Apartment 9D, New York, NY 10028 (US). <b>(74) Agent:</b> CREASON, Gary, L.; Fish & Richardson, P.C., 225 Franklin Street, Boston, MA 02110-02804 (US).		<b>(81) Designated States:</b> JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> ASSAY OF GENE EXPRESSION PATTERNS BY MULTI-FLUOR FISH		
<b>(57) Abstract</b>  <p>An in situ hybridization method for detecting and specifically identifying transcription of a multiplicity of different target sequences in a cell is disclosed. The method includes assigning a different bar code to at least five target sequences, with each target sequence containing at least one predetermined subsequence. Each bar code contains at least one fluorochrome, and at least one bar code comprises at least two different, spectrally distinguishable fluorochromes. A probe set specific for each target sequence is provided in the method. Each probe set contains a hybridization probe complementary to each subsequence in the target sequence. Each probe is labeled with a fluorochrome, and the fluorochromes in each probe set collectively correspond to the bar code for the target sequence of that probe set.</p>		

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<b>AL</b>	Albania	<b>ES</b>	Spain	<b>LS</b>	Lesotho	<b>SI</b>	Slovenia
<b>AM</b>	Armenia	<b>FI</b>	Finland	<b>LT</b>	Lithuania	<b>SK</b>	Slovakia
<b>AT</b>	Austria	<b>FR</b>	France	<b>LU</b>	Luxembourg	<b>SN</b>	Senegal
<b>AU</b>	Australia	<b>GA</b>	Gabon	<b>LV</b>	Latvia	<b>SZ</b>	Swaziland
<b>AZ</b>	Azerbaijan	<b>GB</b>	United Kingdom	<b>MC</b>	Monaco	<b>TD</b>	Chad
<b>BA</b>	Bosnia and Herzegovina	<b>GE</b>	Georgia	<b>MD</b>	Republic of Moldova	<b>TG</b>	Togo
<b>BB</b>	Barbados	<b>GH</b>	Ghana	<b>MG</b>	Madagascar	<b>TJ</b>	Tajikistan
<b>BE</b>	Belgium	<b>GN</b>	Guinea	<b>MK</b>	The former Yugoslav Republic of Macedonia	<b>TM</b>	Turkmenistan
<b>BF</b>	Burkina Faso	<b>GR</b>	Greece			<b>TR</b>	Turkey
<b>BG</b>	Bulgaria	<b>HU</b>	Hungary	<b>ML</b>	Mali	<b>TT</b>	Trinidad and Tobago
<b>BJ</b>	Benin	<b>IE</b>	Ireland	<b>MN</b>	Mongolia	<b>UA</b>	Ukraine
<b>BR</b>	Brazil	<b>IL</b>	Israel	<b>MR</b>	Mauritania	<b>UG</b>	Uganda
<b>BY</b>	Belarus	<b>IS</b>	Iceland	<b>MW</b>	Malawi	<b>US</b>	United States of America
<b>CA</b>	Canada	<b>IT</b>	Italy	<b>MX</b>	Mexico	<b>UZ</b>	Uzbekistan
<b>CF</b>	Central African Republic	<b>JP</b>	Japan	<b>NE</b>	Niger	<b>VN</b>	Viet Nam
<b>CG</b>	Congo	<b>KE</b>	Kenya	<b>NL</b>	Netherlands	<b>YU</b>	Yugoslavia
<b>CH</b>	Switzerland	<b>KG</b>	Kyrgyzstan	<b>NO</b>	Norway	<b>ZW</b>	Zimbabwe
<b>CI</b>	Côte d'Ivoire	<b>KP</b>	Democratic People's Republic of Korea	<b>NZ</b>	New Zealand		
<b>CM</b>	Cameroon		Republic of Korea	<b>PL</b>	Poland		
<b>CN</b>	China	<b>KR</b>	Republic of Korea	<b>PT</b>	Portugal		
<b>CU</b>	Cuba	<b>KZ</b>	Kazakstan	<b>RO</b>	Romania		
<b>CZ</b>	Czech Republic	<b>LC</b>	Saint Lucia	<b>RU</b>	Russian Federation		
<b>DE</b>	Germany	<b>LI</b>	Liechtenstein	<b>SD</b>	Sudan		
<b>DK</b>	Denmark	<b>LK</b>	Sri Lanka	<b>SE</b>	Sweden		
<b>EE</b>	Estonia	<b>LR</b>	Liberia	<b>SG</b>	Singapore		

## **ASSAY OF GENE EXPRESSION PATTERNS BY MULTI-FLUOR FISH**

### **Cross Reference to Related Application**

This application claims benefit of provisional application Serial No. 60/130,563,  
5 filed April 22, 1999.

### **Statement as to Federally Sponsored Research**

Work on this invention was supported by NIH Grant No. GM 54887. Therefore,  
the federal government has certain rights in the invention.

10

### **Background of the Invention**

Recent research on differential gene expression has compared overall gene  
expression in cancer cells with overall gene expression in normal counterpart cells  
(Zhang et al., 1998, *Science* **276**:1268-1272). Similarly, overall gene expression has been  
15 compared in cells undergoing different developmental programs (Chu et al., 1998,  
*Science* **282**:699-705). In such studies, it has been found that large numbers of genes are  
differentially expressed. For example, more than 500 transcripts are expressed at  
significantly different levels in cancer cells versus normal cells. In the case of cancer  
cells, it will be important to correlate the sequences identified as differentially expressed  
20 with actual events occurring at the cellular level or tissue level.

### **Summary of the Invention**

The invention provides an in situ hybridization method for detecting and  
specifically identifying transcription of a multiplicity of different target sequences in a  
25 cell. The method includes assigning a different bar code to at least five target sequences,  
with each target sequence containing at least one predetermined subsequence. Each bar  
code contains at least one fluorochrome, and at least one bar code comprises at least two  
different, spectrally distinguishable fluorochromes. A probe set specific for each target  
sequence is provided in the method. Each probe set contains a hybridization probe  
30 complementary to each subsequence in the target sequence. Each probe is labeled with a  
fluorochrome, and the fluorochromes in each probe set collectively correspond to the bar

code for the target sequence of that probe set. The cell is contacted with a hybridization fluid containing a probe set specific for each target sequence. Following in situ hybridization, fluorochromes on the hybridized probe sets are detected, and spectrally distinguished. This provides separate detection of the transcription site of each target  
5 sequence being expressed. The fluorochromes present at each detected transcription site are related to a bar code, which identifies the target sequence at that transcription site.

Target sequences can include 3 or more, e.g., 4, 5, 6, or 7 predetermined, nonoverlapping subsequences. In some embodiments, at least one target sequence contains subsequences having lengths and spacing between each other so that the  
10 stoichiometry of fluorochromes on probes hybridized with the target sequence is determinable by quantitative fluorescence detection. This can be achieved, for example, by with each subsequence being 30 to 70 nucleotides long, and all the subsequences clustered within a 100-800 nucleotide segment of the target sequence. The region of clustering can be smaller, e.g., 200-600 nucleotides or 300-500 nucleotides. In some  
15 embodiments, each subsequence is about 50 nucleotides long, and all the subsequences are clustered within a 300-nucleotide segment of the target sequence. For maximization of total fluorescence intensity per transcription site, the 100-800 nucleotide segment is located in the 5'-most one third, or 5'-most quarter of the target sequence. Preferably, hybridization probes are labeled with fluorophores attached at intervals of about 5-10  
20 nucleotides. Examples of fluorochromes useful in the invention are Cy2, fluorX, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, fluorescein and Texas red. In some embodiments of the invention, a spectral imaging microscope is employed. Typically the cell is in interphase. The hybridization probe can be an oligonucleotide or a protein nucleic acid (PNA).

25 The invention also provides a probe set panel. The panel contains at least five probe sets, with each probe set being specific for a different target sequence, each of which contains at least one subsequence. Each probe set contains a hybridization probe complementary to each subsequence in the target sequence for which that probe set is specific. Each probe is labeled with a fluorochrome, so that the fluorochromes in each  
30 probe set collectively correspond to a bar code for the target sequence of that probe set.

As used herein, "bar code" means the predetermined, unique combination of fluorochromes assigned to a target sequence.

As used herein, "fluorochrome" means a particular fluorescent dye, e.g., Cy3, without regard to number of individual dye molecules, and without regard to chemical  
5 conjugation.

As used herein, "fluorophore" means an individual fluorescent dye molecule or conjugated moiety.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this  
10 invention belongs. In case of conflict, the present application, including definitions will control. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference.

Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, preferred methods and materials  
15 are described below. The materials, methods, and examples are illustrative only and not intended to be limiting. Other features and advantages of the invention will be apparent from the detailed description and the claims.

### **Brief Description of the Drawings**

20 Fig. 1 is a schematic diagram illustrating the generation of 15 qualitative bar codes using four spectrally distinguishable fluorochromes.

Fig. 2 is a table showing the number of bar codes (distinguishable targets) possible (*Tqual*) when the number of fluorochromes is from 1 to 9 fluorochromes, and qualitative detection is used. The formula for *Tqual* is also shown.

25 Fig. 3 is a schematic diagram illustrating the generation of 27 quantitative bar codes using four spectrally distinguishable fluorochromes and a maximum of three probes (and subsequences) per target sequence.

Fig. 4 is a table showing the number of bar codes (distinguishable targets) possible (*Tquant*) when the number of fluorochromes is from 1 to 7 fluorochromes, the  
30 maximum number of probes (and subsequences) per target sequence is three. The formula for *Tquant* is also shown.

Fig. 5 is a graph showing the emission spectra of seven fluorochromes (fluorescent dyes) that can be used in the invention. The emission spectra shown are those of Cy2, FluorX, Cy3, Cy3.5, Cy5, Cy5.5, and Cy7.

Fig. 6 is a schematic diagram illustrating a voxel and its relationship to nuclear  
5 volume.

### **Detailed Description**

The invention provides a method for specifically assessing the expression of a large number of genes at one time point, in a single cell. This is achieved by  
10 simultaneous in situ hybridization of numerous fluorochrome-labelled probes. Through identification codes based on fluorescence color combinations ("bar codes"), the method allows identification of a large number of target sequences in an interphase nucleus, while using only a small number of spectrally distinguishable fluorochrome labels. For example, by using up to five different hybridization probes per target sequence, where  
15 each probe is labeled with one of five different fluorochromes, the expression of 31 different genes can be detected and identified, using a qualitative fluorescence detection system, i.e., one not capable of determining fluorochrome stoichiometry. When a quantitative fluorescence detection system is employed, i.e., one that determines fluorochrome stoichiometry, expression of up to 232 different target sequences (genes)  
20 can be detected and identified, when only five fluorochromes are employed, with up to five probes per target sequence. The invention can be used to assay the level of gene expression (transcription), ranging from no expression to very high-level expression.

### **Bar Codes and Fluorochromes**

25 Distinguishing and identifying hundreds of different transcription sites in a single nucleus is achieved through the use of a relatively small number of spectrally distinguishable fluorochromes. Each hybridization probe is specific for a subsequence within a target sequence, and each probe is labeled with a fluorochrome. A probe set is designed for each target sequence. A probe set can consist of a single probe. Following  
30 in situ hybridization, the transcription site of each target sequence "lights up" (fluoresces) with a predetermined, unique combination of colors.

Unlike the familiar black and white bar codes read by laser scanners, the colors (fluorochromes) in the "bar codes" of the invention need not appear in any particular order relative to one another. In other words, the encoded information resides in the combinations of colors (and optionally in ratios of color intensity), not in the sequence of  
5 the colors.

In some embodiments of the invention, the detection of fluorochromes on probes hybridized to target sequences is qualitative, i.e., not quantitative. Fig. 1 schematically illustrates how 15 different bar codes can be generated, and thus 15 different targets can be distinguished, when 4 fluorochromes are used. Fig. 2 shows the formula for *Tqual*,  
10 i.e., the number of target sequences that can be identified when using a given number of spectrally distinguishable fluorochromes and qualitative detection. Also in Fig. 2 is a table of values generated using the *Tqual* formula. The table in Fig. 2 shows, for example, that when using 5 fluorochromes and qualitative detection, up to 31 target sequences can be distinguished.

15 In some embodiments of the invention, detection of fluorochromes on probes hybridized to target sequences is quantitative. Fig. 3 schematically illustrates how 12 additional bar codes can be generated, giving a total of 27 different bar codes, when quantitative detection is employed with the same 4 fluorochromes. Fig. 4 shows the formula for *Tquant*, i.e., the number of target sequences that can be identified when using  
20 a given number of spectrally distinguishable fluorochromes with quantitative detection. The *Tquant* formula is written to remove redundant stoichiometries. For example, 2:2 is not used, because it is a multiple of 1:1. Also in Fig. 4 is a table of values generated using the *Tquant* formula. Fig. 4 shows, for example, that when using 5 fluorochromes and quantitative detection, up to 31 target sequences can be distinguished.

25 The difference between qualitative detection and quantitative detection in the invention is illustrated by the following example. With qualitative detection, a target sequence displaying one red fluorochrome and one green fluorochrome would not be distinguishable from a target sequence displaying two red fluorochromes and one green fluorochrome. Both would display the combination of red and green. With quantitative  
30 detection, a target sequence displaying one red fluorochrome and one green fluorochrome (red:green stoichiometry = 1:1) would be distinguishable from a target sequence

displaying two red fluorochromes and one green fluorochrome (red:green stoichiometry = 2:1). The targets would be distinguished by differing ratios of red fluorescence intensity to green fluorescence intensity.

There must always be two colors present to determine the ratio between labeled probes. Absolute quantitation alone is not sufficient. For example, there would be no distinction between a transcription site with two red-labeled probes hybridized to each nascent RNA chain, and a site with twice as many nascent RNA chains bearing only one labeled probe per chain. Similarly, a stoichiometry of 1:1 would not be distinguishable from 2:2, 1:2 would not be distinguishable from 2:4, etc. In general, if the quantitative detection allows differentiation among 1 to 3 probes of each color, the number of distinguishable target sequences increases exponentially (Figs. 2 and 4).

If subsequences (sites where fluorochromes are bound through probe hybridization) are widely spaced within the target sequence, a fluorochrome intensity ratio observed at a transcription site may not accurately reflect the quantitative bar code for that target sequence. For example, a single red-labeled probe hybridizing near the 5' end of the target sequence, and a single green-labeled probe hybridizing near the 3' end of the target sequence, might yield a red:green ratio of 2:1. In each nascent chain, the 5'-located subsequence at which the red-labeled probe binds is transcribed significantly sooner than the 3'-located subsequence at which the green-labeled probe binds. A consequence of this temporal effect can be many more, e.g., twice as many, red-labeled probes bound, in comparison with green-labeled probes, even when the bar code for that target sequence has a red:green stoichiometry of 1:1.

This probe location effect (temporal effect) can be minimized by clustering all of the subsequences within one region of the target sequence. Therefore, in quantitative detection-type embodiments of the invention, it is preferable to use clustered subsequences. In other words, it is preferable to minimize the distance between subsequences. When five 50-nucleotide probes are clustered within a region of 250, 300 or even 400 nucleotides in a target sequence, the fluorochrome stoichiometry in the bar code will predominate over location effects, so that up to 232 target sequences will be distinguishable. Even in qualitative detection-type embodiments it may be advantageous to use clustered subsequences in order to obtain similar signal intensities for each of the



fluorochromes at a given target sequence, i.e., in a given bar code. For maximum total fluorescence signal intensity per transcription site, subsequences are clustered in the 5' portion of the target sequence.

A wide variety of fluorochromes can be used in the invention. The choice of  
5 fluorochromes is not critical, as long as: (1) their emission spectra fall within a suitable range of wavelengths (typically about 350 to 850 nm); and (2) they can be spectrally distinguished from each other by the imaging system with which they will be used. Specific examples of commercially available, spectrally distinguishable fluorochromes useful in the invention are Cy2, FluorX, Cy3, Cy3.5, Cy5, Cy5.5 and Cy7 (CyDyes  
10 Fluorescent Dyes; Amersham Pharmacia Biotech; Piscataway, NJ). Emission spectra of these exemplary fluorochromes are shown in Fig. 5. Other examples of fluorochromes useful in the invention are 6-FAM, HEX, NED, ROX, R110, JOE and TAMRA583.

### **Probe Design and Synthesis**

15 The basic principles of oligonucleotide probe design and synthesis are known in the art. Those basic principles apply generally to the design and synthesis of oligonucleotide probes used in this invention. Similarly, the basic principles of protein nucleic acid (PNA) probe design and synthesis are known in the art, and apply to probes used in this invention.

20 Each hybridization probe is designed to hybridize with a subsequence of a target sequence. Preferably, the sequence complementarity between each probe and the corresponding subsequence is 100%, but this is not required. In some embodiments, useful sensitivity and specificity may be obtained with less than 100% complementarity, e.g., 90%. Preferably, the G/C content of all the probes are matched so the probes have  
25 similar thermal stabilities. There is wide latitude in the choice of hybridization probe length. Considerations in the choice of probe length include total length of target sequence, number of different probes to be accommodated along target sequence, position effect on observed probe stoichiometry, spurious background fluorescence, and fluorescence detection sensitivity. In some embodiments, e.g., embodiments with five  
30 nonoverlapping probes hybridized in tandem within a single 250-500 nucleotide region of the target sequence, a suitable probe length is about 50 nucleotides. In some

embodiments, different probes within a probe set will have the same length and the same number of fluorochrome moieties per probe molecule, but such uniformity is not required. Preferably, within a subsequence cluster the subsequences are separated from each other by a short stretch of nucleotides, e.g. 10 to about 50 nucleotides.

5       Methods of synthesizing DNA generally, including oligonucleotide probes useful in this invention, are known in the art. For a general discussion of oligonucleotide synthesis, see Caruthers, "Synthesis of Oligonucleotides and Oligonucleotide Analogs," in *Topics in Molecular and Structural Biology, Vol. 12: Oligodeoxynucleotides* (Cohen, ed.), MacMillan Press, London, pp. 9-24. Machines for automated DNA synthesis are  
10 commercially available and can be employed in making suitable probes.

Typically, an oligonucleotide probe used in this invention is obtained in a two step process. The first step is synthesis of an oligonucleotide containing a modified base at each position in the nucleotide sequence where a fluorophore is desired. The second step is covalent attachment of a fluorophore to each modified base.

15       The modified base provides a functional group through which the fluorophore is covalently attached to the oligonucleotide. The functional group on the modified base can be any suitable functional group, e.g., a primary amino group. The functional group can be located at the end of a spacer arm.

During the synthesis of the oligonucleotide, the functional group on the modified  
20 base typically bears a protecting group, e.g., a trifluoroacetamide group. Persons skilled in the art will recognize that the protecting group must be removed by a suitable chemical reaction before the functional group can be used for attachment of a fluorophore.

For preparation of amino modified bases, see, e.g., Jablonski et al. *Nucleic Acids Res.* **14**:6115-6128 (1986) and Ruth, *DNA* **3**:123 (1984). A suitable modified base is  
25 incorporated into a deoxythymidine analog used in automated DNA synthesis, and commercially as "Amino-Modifier C6 dT" (Glen Research, Sterling, VA). The total number, and the spacing, of fluorophores on the probe can vary. A preferred spacing is one fluorophore every 5-10 bases. For conventional methods of attaching fluorophores onto amino groups, see Agrawal et al. *Nucleic Acids Res.* **14**:6227-6245 (1986).

30

### **In Situ Hybridization**

In situ hybridization methods useful in the invention are well known in the art. See, e.g., Singer et al., 1986, "Optimization of in situ hybridization using isotopic and nonisotopic detection methods," *Biotechniques* 4:230-250; Lawrence et al., 1989,

- 5 "Highly localized tracks of specific transcripts within interphase nuclei visualized by in situ hybridization," *Cell* 57:493-502; Kislauskis et al., 1993, *J. Cell Biol.* 123:165; and U.S. Patent No. 5,866,331.

In some embodiments, a single probe is labeled with more than one fluorochrome. For example, a single, relatively long probe labeled with two fluorochromes can replace  
10 two relatively short, adjacently-hybridizing probes each labeled with a single fluorochrome.

### **Image Acquisition and Processing**

Persons skilled in the art can carry out the invention using commercially available  
15 fluorescence microscopy instrumentation, i.e., optics, digital imaging hardware, computer hardware and computer software. An imaging technology preferred for sensitive, quantitative detection of fluorochromes is described in Femino et al., 1998, "Visualization of Single RNA Transcripts in Situ," *Science* 280:585-590; and in Singer et al., U.S. Patent No. 5,866,331.

20 Although the imaging technology employed in some embodiments is capable of visualizing a single mRNA molecule, single molecule sensitivity is not required. That is because numerous nascent RNA transcripts are produced at any actively expressing gene. The nascent RNA transcripts provide highly localized amplification of the target sequence. The degree of gene expression is indicated by total fluorescence intensity at  
25 the site. The total fluorescence intensity is positively correlated with the number of hybridized probes, which depends on the number of nascent RNA chains.

The spectral characteristics of transcription sites subjected to in situ hybridization involving fluorochrome bar codes can be determined using any suitable technology. One useful approach to spectral characterization is based on conventional optical bandpass  
30 filters. Another approach (discussed below) is spectral imaging microscopy.

The invention advantageously exploits the fact that a gene transcription site occupies a very small volume, i.e., approximately of 5 voxels, wherein:

$$1 \text{ voxel} = 0.1 \text{ } \mu\text{m} \times 0.1 \text{ } \mu\text{m} \times 0.25 \text{ } \mu\text{m} = 0.0025 \text{ } \mu\text{m}^3$$

A 10  $\mu\text{m}$  diameter nucleus contains a volume of approximately  $500 \text{ } \mu\text{m}^3$ , i.e.,  $\frac{4}{3} \pi r^3$ . Thus, the volume of one nucleus is equivalent to approximately 200,000 voxels. One transcription site occupies approximately 1/40,000 of the nuclear volume (Fig. 6). Thus, each transcription site has a distinct location. If two transcription sites are very close, a super-resolution algorithm (Carrington et al., 1995, *Science* **268**:1483-1487) which defines the center of density of the fluorescence can compress the transcription  
10 into one voxel. This would allow as many as 40,000 transcription sites to be mapped within the nucleus. The spatial resolution of such technology is able to distinguish the expression of every gene in a human cell. This level of spatial resolution is unlikely to be needed, because an individual cell is likely to be expressing no more than 5,000 genes at any moment.

15 Another imaging technology useful in the invention is spectral imaging microscopy (Schrock et al, 1996, *Science* **273**:494-497). This technology combines Fourier spectroscopy, charge-coupled device (CCD) imaging, and optical microscopy to measure simultaneously at all points in the sample emission spectra. The spectral imaging microscope (Applied Spectral Imaging, Inc.; Carlsbad, CA) is capable of  
20 interrogating the spectral characteristics of each individual pixel in a two-dimensional image.

The 3-dimensional algorithms used in visualization of single RNA transcripts in situ (Femino et al., 1998, *Science* **280**:585-590) can be used with spectral imaging microscopy to analyze the spectral character of each transcription site in a 3-dimensional  
25 series of images. This approach advantageously distinguishes single discrete spectral signals from the autofluorescence background present in cells and tissue samples. In addition, it achieves high spectral separation, i.e., 15 nm resolution. This increases the number of fluorochromes that can be employed at one time. In addition, the spectral imaging microscope permits greater collection of light output per fluorochrome, as  
30 compared to barrier filters, which require a narrow band pass to distinguish signals.

### **Uses of the Invention**

In general, the invention is useful for clinical diagnostic testing and as a research tool in cell biology and molecular genetics. For example, it can be used to diagnose cancer by identifying genes that are differentially expressed in cancer cells, as compared to normal control cells. In addition, a probe set for a particular panel of genes can be used to identify a particular type of cancer, to ascertain the cancer's tissue of origin, to predict whether it will be metastatic, or to predict the susceptibility of the cancer to certain drugs. Similarly, a probe set for a genetic disease can be used to determine if a particular genetic disease is present. A probe set for an infectious disease can be used to detect activity of a virus, e.g., HIV or herpes. Probe sets can be used to detect gene expression indicative of hormonal imbalance. In an example of application as a research tool, the invention is useful to confirm, at the single cell level, differential sequence expression data obtained by testing extracted RNA on DNA microarrays (chips).

The invention is further illustrated by the following examples. The examples are provided for illustration purposes only. They are not to be construed as limiting the scope or content of the invention in any way.

### **Examples**

#### **Example 1: Qualitative Detection**

Demonstration of a qualitative detection-type embodiment of the invention is carried out using cells from cultured colorectal (CR) cell lines derived from a human colorectal tumor. These cells are chosen for reasons including the following: They provide an optically optimal model. The nucleus is somewhat flattened in cells spread in culture, but it provides good X-Y spatial discrimination. Considerable molecular genetic characterization is available from previous work on these cells. A high similarity between gene expression profiles of some CR cultured cell lines and CR primary tissue is reported to exist (Zhang et al., 1997, *Science* **276**:1268-1272). For example, about half the genes identified by SAGE (serial analysis of gene expression) as overexpressed in primary tumors relative to normal tissue are reported to be increased also in certain colorectal cell lines.

In a first series of experiments, 31 target sequences are chosen, from among sequences previously identified by SAGE analysis as overexpressed in colorectal cell lines relative to normal control cells. Twenty target sequences are chosen because they are thought to exhibit a binary expression pattern, i.e., ten that are off in tumor cells, and  
5 on in normal cells, and ten that are on in tumor cells, and off in normal cells. Eleven target sequences are chosen because they are thought to be constitutive in cancer cells and in normal cells.

Using a total of five spectrally distinguishable fluorochromes, 31 different bar codes are created without using a given fluorochrome more than once in a given bar code.  
10 The creation of the 31 bar codes using 5 fluorochromes is an extension of the scheme illustrated in Fig. 1, where 15 qualitative bar codes are created using 4 fluorochromes. One of the 31 bar codes is assigned to each of the 31 target sequences.

From one to five 50-nucleotide subsequences clustered near the 5' end of each target sequence are chosen. For each subsequence, a complementary 50-nucleotide probe  
15 is chemically synthesized, using an automated DNA synthesizer and commercial reagents, in accordance with the vendors' instructions. Each probe contains five amino-modified bases spaced at ten-nucleotide intervals.

A commercially available cyanine fluorochrome is covalently attached to the amino-modified bases on each probe according to the fluorochrome vendor's instructions.  
20 The fluorochrome to be attached to each probe is chosen so that the set of 5 probes for each target sequence collectively contains exactly the same fluorochrome(s) combination present in the bar code assigned to that target sequence.

Fixation of interphase cells and FISH protocol are essentially as described in Femino et al. (*Science* **280**:585-590), except that the hybridization fluid contains the  
25 fluorochrome-labeled hybridization probes for all of the subsequences in all 31 target sequences.

Following completion of the FISH protocol, interphase nuclei are subjected to fluorescence microscopy, using optical filters essentially as described by Speicher et al., 1996, *Nature Genetics* **12**:368-375. In almost all nuclei examined, 21 foci, or elongated  
30 tracks, of fluorescence are detected. In some nuclei, two foci appear to overlap. In a few cases, the overlap is too great to allow spectral resolution of the two foci. Although the

dimensions and intensities of the foci and tracks vary considerably, it appears that even the smallest and weakest focus or track represents the combined signals from numerous nascent RNA chains at a transcription site. In most nuclei examined, spectral analysis permits the association of each focus or track with one of the 31 bar codes, at a high  
5 confidence level.

### **Example 2: Quantitative Detection**

A second series of experiments is carried out. In the second series, five spectrally distinguishable fluorochromes are used to create 232 different bar codes. Many of the  
10 bar codes are distinguished from other bar codes only when fluorochrome ratios within combinations are considered. This creation of the 232 bar codes using 5 fluorochromes is an extension of the scheme illustrated in Fig. 3, where 27 quantitative bar codes are created using 4 fluorochromes. One of the 232 bar codes is assigned to each of 232 target sequences.

15 Cellular material, subsequence selection, probe synthesis, cell fixation, and FISH are carried out as in the first series of experiments described above.

Following completion of the FISH protocol, interphase nuclei are subjected to fluorescence microscopy. The fluorescence microscopy used in this series of experiments combines the methodology of Femino et al. (*Science* **280**:585-590) and  
20 Singer et al. (U.S. Pat. No. 5,866,331) with the spectral imaging methodology of Schrock et al. (*Science* **273**:494-497). This approach to the acquisition and processing of fluorescence information from the hybridized probes provides spectral resolution and signal quantitation sufficient for ascertaining fluorochrome stoichiometry at most transcription sites. In addition, this approach provides 3-dimensional, spatial resolution  
25 far greater than that required to resolve 232 transcription sites from each other within the nuclear volume.

### **Example 3: Correlation of FISH Data with DNA Microarray Data**

30 Data from microarray analysis of RNA isolated from cultured colon cancer cell lines and normal control lines are used to select 31 differentially expressed target

sequences. Subsequences are selected, qualitative bar codes are created, labeled probes are synthesized, in situ hybridization is carried out, and fluorescence microscopy is performed, all essentially as described in Example 1 (above). FISH data are used to characterize relative expression levels of the 31 target sequences.

5       The FISH-derived, relative expression level of each target sequence is compared to the microarray-derived, relative expression level of the same sequence. Even though the identity of the expression product of many of the target sequences is not known, a positive correlation of the FISH data and the microarray data is established.

10       The comparison shows that a large proportion of sequences indicated by microarray analysis to have a certain expression level in a certain cell type are consistently expressed at the indicated level from cell to cell. The comparison also shows, however, that some sequences indicated by microarray analysis to have a certain expression level are actually expressed at widely varying levels from cell to cell. Such variably-expressed sequences are identified as having relatively little predictive or  
15       diagnostic value.

#### **Example 4: Simultaneous Bar Code Detection of Eight Genes in Cultured Cells**

20       Cultured human colon cancer cells (line DLD-1) and cultured human foreskin fibroblast cells were used. Cells were induced with a serum pulse and cycloheximide before preparation, to enhance the transcription sites. Probe sets, i.e., 3 non-overlapping, antisense probes for each target gene, were synthesized using conventional methods and a commercial DNA synthesizer. For subsequent attachment of fluorophores, commercial, modified nucleotides (amino modified C6-DT; Glenn Research, Sterling, VA) were  
25       incorporated during probe synthesis. Each probe was 50 nucleotides in length, and was labeled with five fluorophores.

30       Each 50-nucleotide probe molecule was labeled with a single type of fluorochrome. For example, the probe set for the  $\beta$ -actin gene or gene transcript (bar code: FITC, cy3, cy5) included the following: (1) a 50-nucleotide probe complementary to a first 50-nucleotide subsequence in the  $\beta$ -actin transcript, and labeled with five FITC fluorophores; (2) a 50-nucleotide probe complementary to a second 50-nucleotide



subsequence in the  $\beta$ -actin transcript, and labeled with five cy3 fluorophores; and (3) a 50-nucleotide probe complementary to a third 50-nucleotide subsequence in the  $\beta$ -actin transcript, and labeled with five cy5 fluorophores. The following fluorochromes were used in varying combinations to label the probes and form bar codes: FITC, cy3, cy5,  
 5 and cy7 (Table 1).

**Table 1**

	<b><u>Target Gene</u></b>	<b><u>Bar Code</u></b>
	$\beta$ -actin	FITC, cy3, cy5
10	c-myc	FITC, FITC, cy3
	cyclin D1	FITC, FITC, cy5
	gamma-actin	FITC, FITC, cy7
	c-jun	cy3, cy3, cy5
	SRF (serum response factor)	cy3, cy3, cy7
15	COX-2 (cytochrome C oxidase)	cy5, cy5, cy7
	c-fos	cy3, cy5, cy7

Cultured cells grown on cover slips and treated with TRITON X-100® and fixed with 4% paraformaldehyde for 30 minutes and washed in phosphate buffered saline  
 20 (PBS). Probe hybridization (30 ng of DNA in each probe set) was carried out under standard conditions, i.e., 50% formamide, 2xSSC, 37°C, for 3 hours. After removal of non-hybridized probes by washing (50% formamide, 2xSSC), cells were mounted on slides in glycerol/phenylenediamine containing DAPI, for viewing by fluorescence microscopy. Imaging of cells was carried out using 60X objective magnification and a  
 25 CCD camera (Photometrics). Appropriate filter sets were used for each fluorochrome. Images from each filter set were pseudocolored and merged to that one image could be analyzed for overlapping colors at transcription sites. Nuclei were identified by DAPI signal.

Transcription sites of target genes were identified by setting a detection threshold to eliminate weak signals. The transcription sites were the brightest signals in the field. Positive confirmation was determined by the presence of at least two colors at a particular site. A computer programs was written to identify the transcription sites by interrogating  
5 bright regions of signal with a prescribed size and pseudocolor identification. The program detected and distinguished transcription at each of the eight target genes within a single sample preparation of colon cancer cells, solely by fluorescence barcode data. In the fibroblasts, similar results were obtained.

10 Other embodiments are within the following claims.

**WHAT IS CLAIMED IS:**

1           1. An in situ hybridization method for detecting and specifically identifying  
2 transcription of a multiplicity of different target sequences in a cell, the method comprising:  
3           assigning a different bar code to at least five target sequences, each target sequence  
4 comprising at least one predetermined subsequence, wherein each bar code comprises at least  
5 one fluorochrome, and at least one bar code comprises at least two different, spectrally  
6 distinguishable fluorochromes;  
7           providing a probe set specific for each target sequence, each probe set containing a  
8 hybridization probe complementary to each subsequence in the target sequence, each probe  
9 being labeled with a fluorochrome, wherein the fluorochromes in each probe set collectively  
10 correspond to the bar code for the target sequence of that probe set;  
11           contacting the cell with a hybridization fluid comprising a probe set specific for each  
12 target sequence;  
13           detecting fluorochromes on the probe set hybridized to RNA transcribed from each  
14 target sequence, if present, thereby separately detecting transcription sites, wherein the  
15 detecting includes spectrally distinguishing the different fluorochromes; and  
16           relating the fluorochromes at each detected transcription site to a bar code.

1           2. The method of claim 1, wherein at least one target sequence comprises three or  
2 more predetermined, nonoverlapping subsequences.

1           3. The method of claim 2, wherein at least one target sequence contains  
2 subsequences having lengths and spacing between each other so that the stoichiometry of  
3 fluorochromes on probes hybridized with the target sequence is determinable by quantitative  
4 fluorescence detection.

1           4. The method of claim 3, wherein each subsequence is 30 to 70 nucleotides long,  
2 and all the subsequences are clustered within a 100-800 nucleotide segment of the target  
3 sequence.

1           5. The method of claim 4, wherein the subsequences are clustered within a 200-600  
2 nucleotide segment of the target sequence.

1           6. The method of claim 5, wherein the subsequences are clustered within a 300-500  
2 nucleotide segment of the target sequence.

1           7. The method of claim 6, wherein each subsequence is about 50 nucleotides long,  
2 and all the subsequences are clustered within a 500-nucleotide segment of the target  
3 sequence.

1           8. The method of claim 3, wherein each subsequence is 30 to 70 nucleotides long,  
2 and all the subsequences are clustered within a nucleotide segment that represents about 10%  
3 of the total target sequence length.

1           9. The method of claim 4, wherein the 100-800 nucleotide segment is in the 5'-most  
2 one third of the target sequence.

1           10. The method of claim 9, wherein the 100-800 nucleotide segment is in the 5'-most  
2 quarter of the target sequence.

1           11. The method of claim 1, wherein the hybridization probe is labeled with a  
2 multiplicity of fluorophores attached at intervals of 5-10 nucleotides.

1           12. The method of claim 1, wherein the fluorochromes are selected from the group  
2 consisting of Cy2, fluorX, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, fluorescein and Texas red.

1           13. The method of claim 1, wherein fluorochromes are detected using a spectral  
2 imaging microscope.

1           14. The method of claim 1, wherein the cell is in interphase.

1           15. The method of claim 1, wherein the hybridization probe is an oligonucleotide.

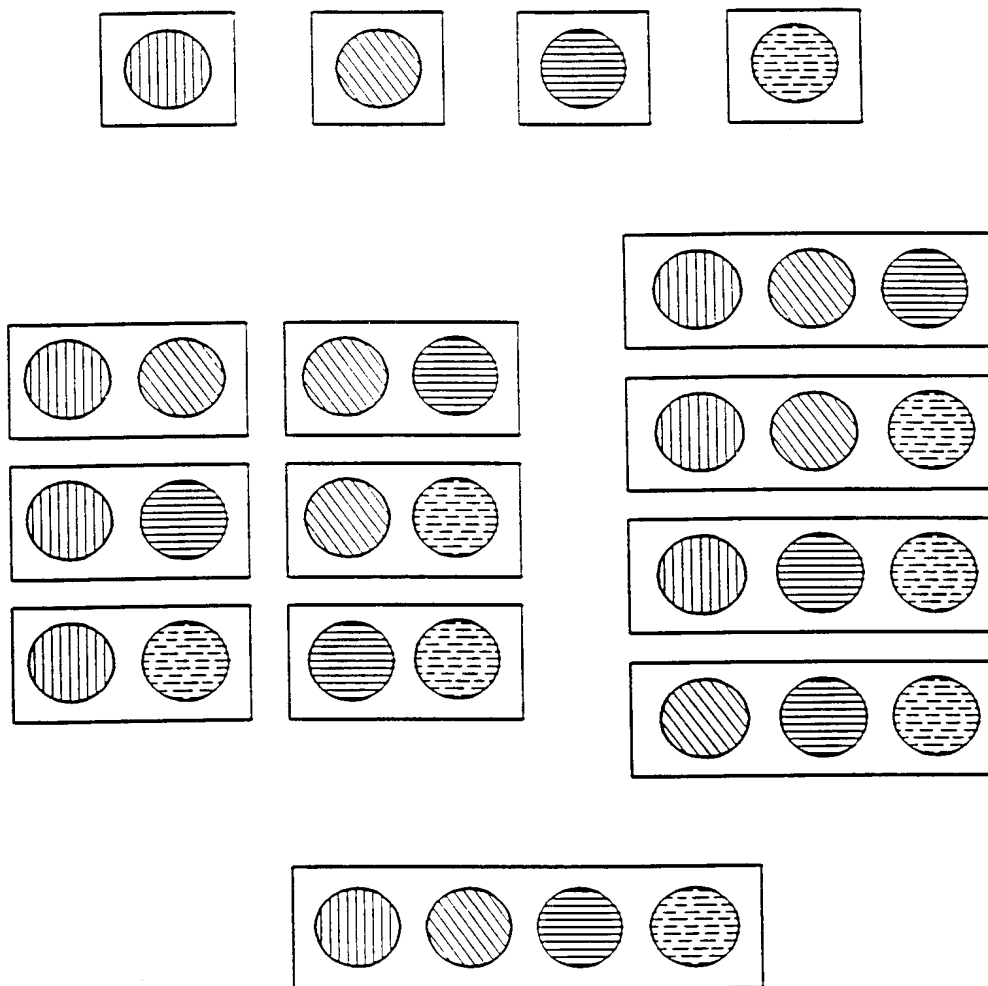
1           16. The method of claim 1, wherein the hybridization probe is a protein nucleic acid  
2 (PNA).

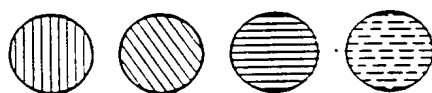
1           17. The method of claim 1, wherein the cell is part of a tissue sample.

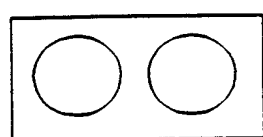
1           18. A probe set panel comprising at least five probe sets, each probe set specific for a  
2 different target sequence, each target sequence comprising at least one subsequence, each  
3 probe set containing a hybridization probe complementary to each subsequence in the target

- 4 sequence for which that probe set is specific, each probe being labeled with a fluorochrome,
- 5 so that the fluorochromes in each probe set collectively correspond to a bar code for the
- 6 target sequence of that probe set.

1/5



 = distinct fluorochromes

 = distinguishable target  
(bar code)

**FIG. 1**

2/5

## Qualitative Detection

$$T_{qual} = n + \left[ \sum_{r=2}^{n-1} \frac{n!}{r!(n-r)!} \right] + 1$$

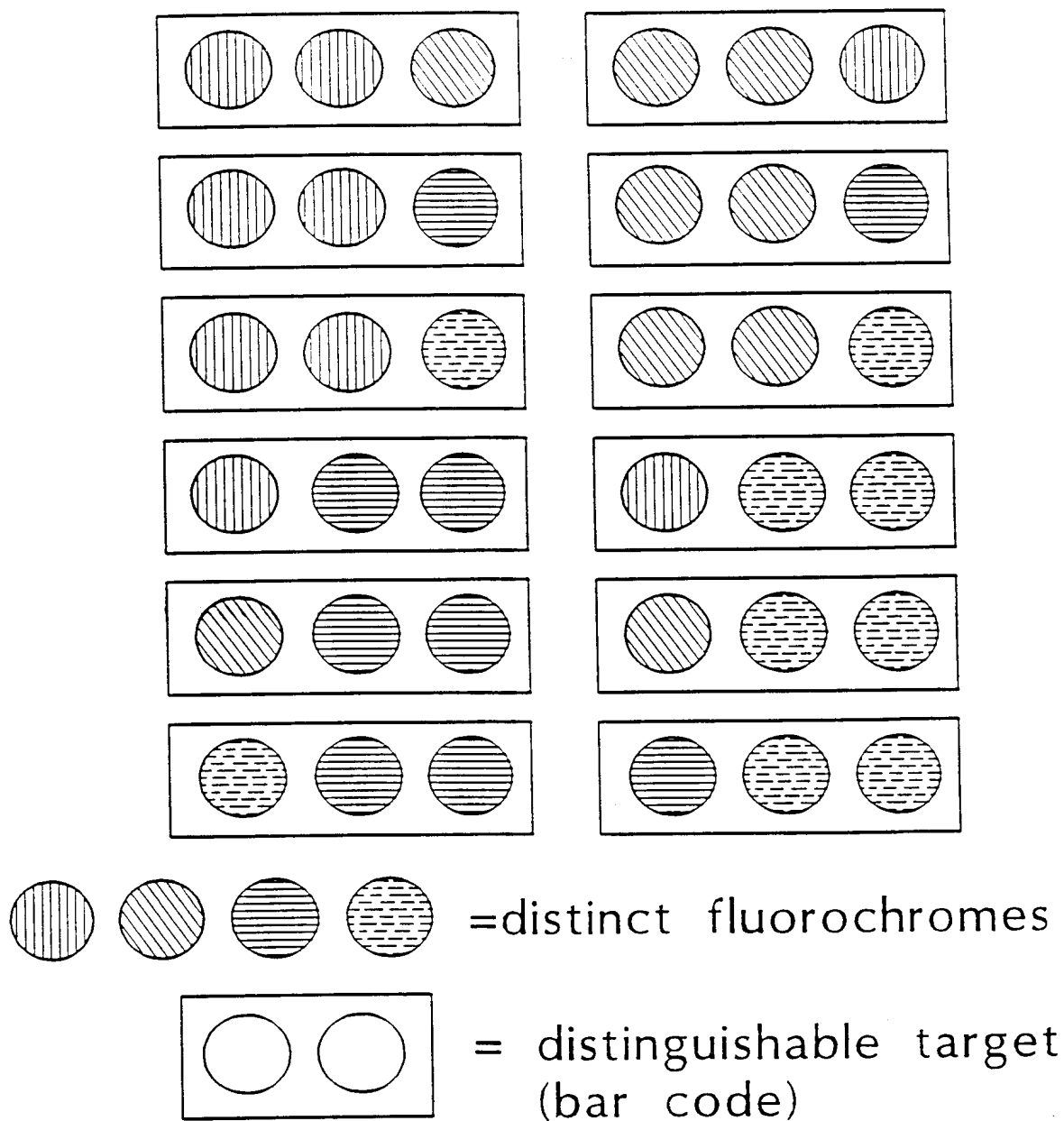
n = number of distinct fluorochromes    T = number of distinguishable targets    number of unique probes

1	1	1
2	3	4
3	7	12
4	15	32
5	31	80
6	63	192
7	127	448
8	255	1024
9	511	2304

**FIG. 2**

3/5

15 qualitative targets +

**FIG. 3**



4/5

## Quantitative Detection

$$T_{quant} = \sum_{r=1}^n \left[ \frac{(m+r-1)!}{r!(m-1)!} - m \right] + n + 1$$

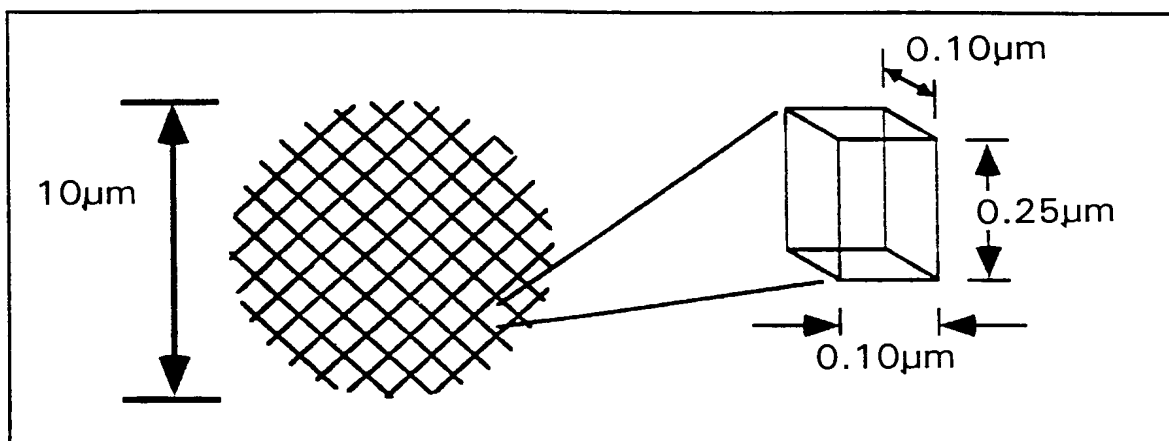
T = number of distinguishable targets using unique stoichiometries of each fluorescently labeled probe (eg 2:2 is not used because it is a multiple of 1:1)

m = maximum number of probes per target

n = number of distinct fluorochromes	T	2	3	4	5	6	7
	2	4	6	8	13	14	24
	3	7	14	23	44	57	102
	4	11	27	52	110	166	306
	5	16	46	101	232	397	762
	6	22	72	177	438	834	1680
	7	29	106	288	764	1597	3390

FIG. 4

5/5



nucleus

voxel

$$V = \frac{4}{3} \pi r^3 = 500 \mu\text{m}^3$$

$$V = 2.5 \times 10^{-3} \mu\text{m}^3$$

FIG. 6

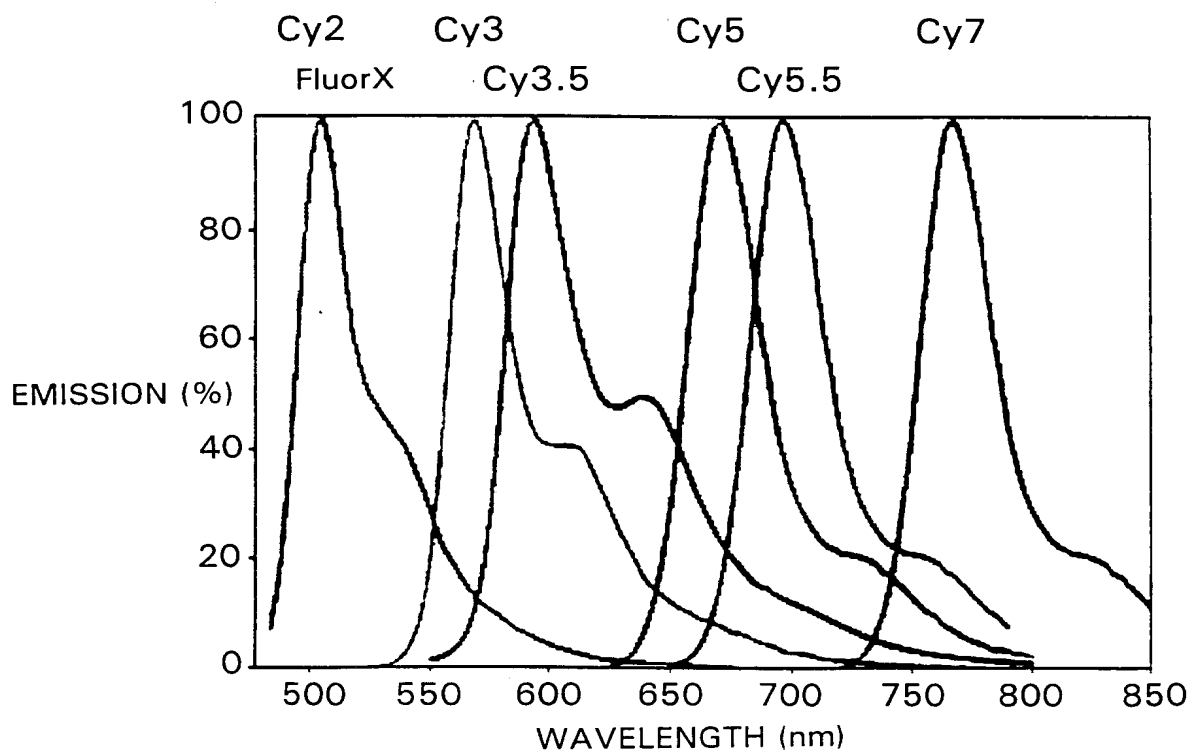


FIG. 5

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
2 November 2000 (02.11.2000)

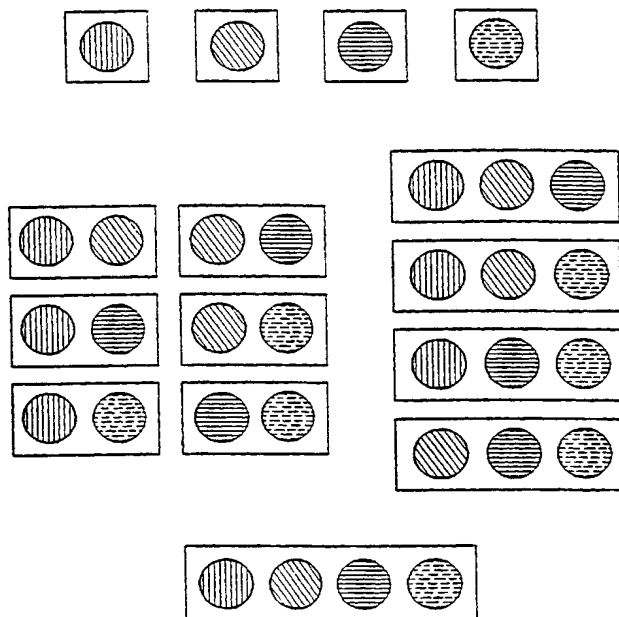
PCT

(10) International Publication Number  
**WO 00/65094 A3**

- (51) International Patent Classification<sup>7</sup>: **C12Q 1/68** **YESHIVA UNIVERSITY** [US/US]; 1300 Morris Park Avenue, Bronx, NY 10461 (US).
- (21) International Application Number: PCT/US00/10879
- (22) International Filing Date: 21 April 2000 (21.04.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/130,563 22 April 1999 (22.04.1999) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/130,563 (CIP)  
Filed on 22 April 1999 (22.04.1999)
- (71) Applicant (for all designated States except US): **THE ALBERT EINSTEIN COLLEGE OF MEDICINE OF**
- (72) Inventor; and
- (73) Inventor/Applicant (for US only): **SINGER, Robert, H.** [US/US]; 144 East 84th Street, Apartment 9D, New York, NY 10028 (US).
- (74) Agent: **CREASON, Gary, L.**; Fish & Richardson, P.C., 225 Franklin Street, Boston, MA 02110-02804 (US).
- (81) Designated States (national): JP, US.
- (84) Designated States (regional): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
- Published:  
— with international search report

[Continued on next page]

(54) Title: ASSAY OF GENE EXPRESSION PATTERNS BY MULTI-FLUOR FISH



(57) Abstract: An in situ hybridization method for detecting and specifically identifying transcription of a multiplicity of different target sequences in a cell is disclosed. The method includes assigning a different bar code to at least five target sequences, with each target sequence containing at least one predetermined subsequence. Each bar code contains at least one fluorochrome, and at least one bar code comprises at least two different, spectrally distinguishable fluorochromes. A probe set specific for each target sequence is provided in the method. Each probe set contains a hybridization probe complementary to each subsequence in the target sequence. Each probe is labeled with a fluorochrome, and the fluorochromes in each probe set collectively correspond to the bar code for the target sequence of that probe set.

= distinct fluorochromes

= distinguishable target (bar code)

WO 00/65094 A3



**(88) Date of publication of the international search report:**  
16 August 2001

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# INTERNATIONAL SEARCH REPORT

Int'l Application No

PCT/US 00/10879

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DAUWERSE ET AL.: "Multiple colors by FISH using ratio labelled DNA probes create a molecular karyotype" HUMAN MOL GEN, vol. 1, no. 8, 1992, pages 593-598, XP002161585 the whole document	1-18
X	WO 97 23648 A (UNIV YALE) 3 July 1997 (1997-07-03) page 8; claims	1,2, 11-18
	--- -/-- ---	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

\* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*&\* document member of the same patent family

Date of the actual completion of the international search

27 February 2001

Date of mailing of the international search report

13/03/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Molina Galan, E

# INTERNATIONAL SEARCH REPORT

Int. :ional Application No

PCT/US 00/10879

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>SCHROECK E ET AL: "MULTICOLOR SPECTRAL KARYOTYPING OF HUMAN CHROMOSOMES" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, ,US, vol. 273, 26 July 1996 (1996-07-26), pages 494-497, XP000952836 ISSN: 0036-8075 cited in the application the whole document</p>	1,2, 11-18
X	<p>SPEICHER M R ET AL: "KARYOTYPING HUMAN CHROMOSOMES BY COMBINATORIAL MULTI-FLUOR FISH" NATURE GENETICS, NEW YORK, NY, US, vol. 12, 12 April 1996 (1996-04-12), pages 368-375, XP000930084 ISSN: 1061-4036 the whole document</p>	1,2, 11-18
A	<p>WO 97 40191 A (BUCKWALD ROBERT A ; CABIB DARIO (IL); GARINI YUVAL (IL); APPLIED SP) 30 October 1997 (1997-10-30)</p>	
A	<p>WO 97 14816 A (UNIV MASSACHUSETTS) 24 April 1997 (1997-04-24) cited in the application page 7, line 19 -page 10, line 14</p>	3-10
A	<p>LENGAUER C ET AL: "CHROMOSOMAL BAR CODES PRODUCED BY MULTICOLOR FLUORESCENCE IN SITU HYBRIDIZATION WITH MULTIPLE YAC CLONES AND WHOLE CHROMOSOME PAINTING PROBES" HUMAN MOLECULAR GENETICS, GB, OXFORD UNIVERSITY PRESS, SURREY, vol. 2, no. 5, 1 May 1993 (1993-05-01), pages 505-512, XP000612071 ISSN: 0964-6906</p>	
A	<p>EP 0 612 851 A (UNIV MASSACHUSETTS MEDICAL) 31 August 1994 (1994-08-31)</p>	

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/10879

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9723648	A	03-07-1997	US 5759781 A	02-06-1998
			AU 724912 B	05-10-2000
			AU 1423597 A	17-07-1997
			CA 2241148 A	03-07-1997
			CA 2241364 A	03-07-1997
			EP 0874912 A	04-11-1998
			EP 0879297 A	25-11-1998
			JP 11510707 T	21-09-1999
			US 6007994 A	28-12-1999
			WO 9723649 A	03-07-1997
			JP 2000507809 T	27-06-2000
WO 9740191	A	30-10-1997	US 5817462 A	06-10-1998
			AU 2730597 A	12-11-1997
			EP 0896631 A	17-02-1999
			JP 2000509977 T	08-08-2000
			US 6066459 A	23-05-2000
			US 5719024 A	17-02-1998
			US 5856871 A	05-01-1999
			US 5835214 A	10-11-1998
			US 5798262 A	25-08-1998
			US 6018587 A	25-01-2000
			US 5912165 A	15-06-1999
			US 5906919 A	25-05-1999
WO 9714816	A	24-04-1997	US 5866331 A	02-02-1999
			AU 7460096 A	07-05-1997
EP 0612851	A	31-08-1994	US 5563033 A	08-10-1996
			US 5955272 A	21-09-1999